A Multiservice Switch for Advanced Avionics Data Networks

W. A. Rosen[1]
W. A. Phipps[2]
A. D. George[2]
D. D. Turner[1]
V. Gershman[1]
M. P. Birmingham[1]

1. Naval Air Warfare Center, Aircraft Division
Code 4551 M/S 07
Warminster, PA 18974
U.S.A

2. High-performance Computing and Simulation Research Lab
FAMU-FSU College of Engineering
2525 Pottsdamer Street
Tallahassee, FL 32310
U.S.A

Abstract

With knowledge of persistent data communication traffic patterns offered to an avionics data network, modifications to the routing through the network can be made to improve total throughput and bound the latency of packets. The Multiservice Switch (MSS) is such a route-optimizing switch for streaming sensor data. The MSS has two switching fabrics: packet switching and circuit switching. The packet-switching fabric routes small control and data packets between switch ports. The circuit-switching fabric uses a crossbar to physically connect ringlets, which reduces the workload on the packet-switching fabric for long data streams between the ports.

An implementation of the MSS is described which uses commercial-off-the-shelf (COTS) components. A simulation model was developed to show the benefits of the MSS under standard avionics workloads. The results of the MSS indicate distinct advantages in terms of performance, price, and power consumption over other conventional switch and network topology designs.

Introduction

A number of recent studies have identified a requirement for a unified avionics data network that is capable of replacing a variety of existing interconnects such as the Parallel Interface (PI) Bus, Data Network/Data Flow Network (DN/DFN), High Speed Data Bus (HSDB), and Sensor Data Distribution Network (SDDN) [UHLH92][SAE93]. For example, studies performed under the Air Force PAVE PACE and Very High Speed Optical Networks (VHSON) programs have shown that by integrating the functionality of the DN/DFN, PI Bus, HSDB, and sensor/video network into a single network, the reliability of the interconnects could increase by a factor of 13 while reducing cost by 50%, weight by 60%, and power by 70% [ULHL92]. As a result, system designs such as the Joint Strike Fighter (JSF) preferred concept feature a unified network as an essential component of the architecture [JAST94].

One of the difficulties impeding the implementation of a unified network is the development of a data switch capable of supporting the conflicting requirements of the networks being replaced. For example, PI Bus traffic is characterized by short, low-latency messages which would best be handled by a connectionless, packet-switched transfer whereas DN/DFN traffic is characterized by stream data best handled by a connection-oriented, circuit-switched network. Sensor data is a mix of the two in that it is mostly stream data interrupted occasionally by very-low-latency, high-integrity control and status information.

In this paper we describe the development of a compact, low-power multiservice switch capable of supporting both connectionless and connection-oriented transfers. The switch operates at a 1-Gbps serial data rate and the inputs and outputs are optical. The switch is based on the IEEE 1596-1992 Scalable Coherent Interface (SCI) standard [SCI93]. This standard supports a number of interconnect topologies including ringlets, switched networks, and ringlets interconnected by switches which make it suitable for multiservice transfers. The MSS provides multiservice support by incorporating a crossbar switch which reconfigurably interconnects ringlets to form larger ringlets. In addition, each input port is connected by a back-end bus which reroutes messages addressed to nodes on other ringlets. Stream data transfers are supported by connecting the source and target nodes on a common ringlet via the crossbar switch, while small, bursty transfers are supported via the back-end bus.

The advantage of this topology is that the back-end bus is only used to transfer relatively short control and status messages, so that very-low latency can be achieved for these messages. An added advantage is that the power, size, and cost of the switch are much lower than in a switch that must provide high-speed, exclusively-connectionless transfers. In the next sections we describe the functional design of the switch and predicted performance and power dissipation for a 5-port (4 SCI ports, 1 control port) prototype currently undergoing test and evaluation. This switch is based on the Dolphin LC-1 link controller chip which uses interval routing. We also describe the results of simulations that predict the performance of a switch based on look-up table routing which would provide greater system flexibility. Finally, brief conclusions are drawn about the performance and utility of the multiservice switch.

SCI Overview

SCI is a unidirectional, point-to-point, high-performance network protocol with a standard bandwidth of 1-GBps and a media access control using register insertion ring for low-latency concurrent transfers. SCI is a synchronous protocol and emits a single 18-bit symbol at each clock cycle. SCI packets are made up of a series of delimited symbols. The internal structure of an SCI node is shown in Figure 1.

Incoming SCI packets arrive and are routed either to the input queue or to the bypass FIFO by the stripper based on the destination address of the packet. The host interface services the input queue and offers new packets into the output queue. A multiplexer arbitrates between the bypass FIFO and output queue for transmission onto the SCI ring.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
1996		**00-00-1996 to 00-00-1996**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
A Multiservice Switch for Advanced Avionics Data Networks	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Naval Air Warfare Center,Aircraft Division,Code 4551,Warminster,PA,18974**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**		**6**	

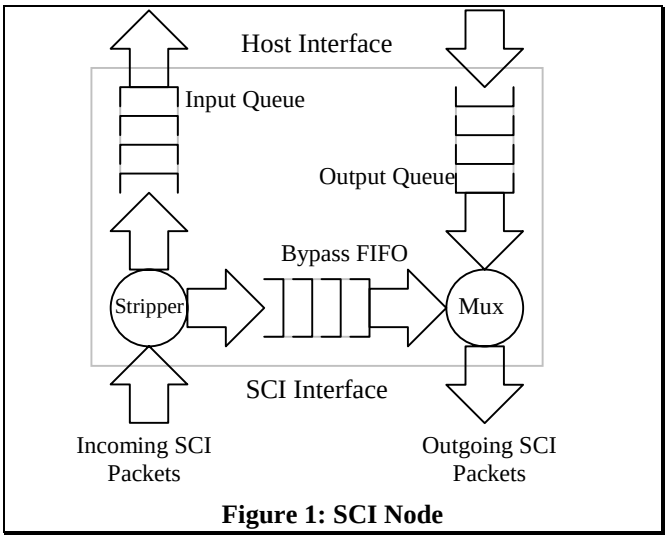

Figure 1: SCI Node

Common SCI topologies are ring-based so that packets are passed through the bypass FIFOs of intermediate nodes on their way to the destination node. Although rings are the easiest topology to create using SCI nodes, they suffer from a lack of fault tolerance and a minimum latency proportional to the number of intermediate nodes. SCI switches are used to connect separate SCI rings in an attempt to increase both fault tolerance as well as improve performance by routing packets out of rings to save bandwidth. Switches have a penalty of routing delay, which is necessary for all packets that are routed by the switch. Certainly a trade-off between the performance improvements of a switch and the streaming performance of the ring can be made.

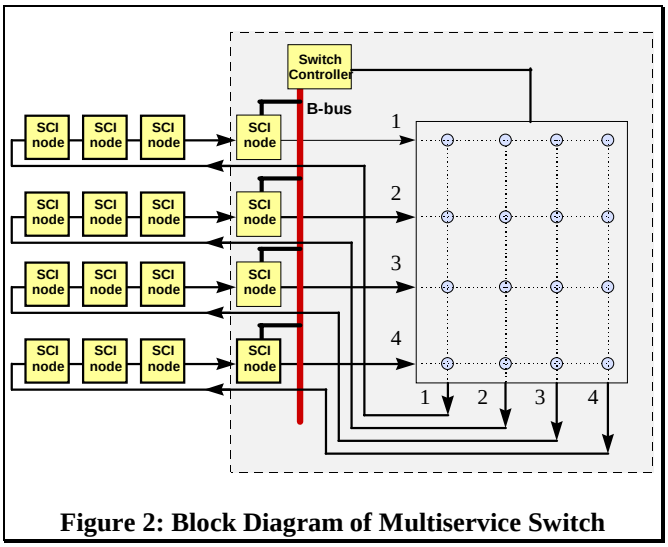

Figure 2: Block Diagram of Multiservice Switch

Switch Design

Figure 2 shows a functional block diagram of the multiservice switch. The default configuration has the crossbar simply passing packets from the same numbered input port to output port. Figure 3 shows a schematic of an individual port inside the switch. Each port on the MSS is connected to an SCI ringlet consisting of several nodes. The serial optical input signal at each port is converted to an electrical signal and inputted to an Hewlett Packard G-Link chip for deserializing and decoding. The parallel format is required for SCI node interface (i.e. the Dolphin LC-1) that receives it next. The output of the LC-1 is encoded, converted back to serial, and sent to one of the inputs of a serial, electronic-crossbar switch. The corresponding output of the crossbar is converted to an optical signal and routed to the output of the port, where it completes the ringlet. The crossbar switch is controlled via a parallel port which may be attached to a host processor connected to any node on the network. The same host controls the initialization and status of the LC-1 chip at each port via separate control logic. The node interfaces at each port are connected together via a back-end bus (i.e. the B-bus in Figure 2). Packets addressed to a ringlet other than the one to which the port is connected are stripped from the ringlet by the interface circuit and routed to the appropriate ringlet via the back-end bus.

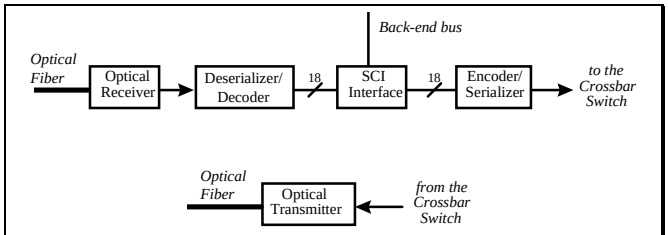

Figure 3: Multiservice Switch Port Schematic

Individual ringlets may be connected together through the crossbar switch to form a single ringlet. For example, if the crossbar switch is configured so that input 1 is connected to output 4 and input 4 is connected to output 1, all of the nodes in ringlets 1 and 4 actually reside on a common ringlet. A typical configuration might consist of a sensor on one ringlet connected to a second ringlet comprised of a suite of processing and memory modules. Stream data from the sensor is transferred to the processing suite through the crossbar switch. Short control and status messages from or to nodes residing on different ringlets are transferred over the back-end bus. Since only the low-data-rate control and status messages are transferred over the back-end bus, very-low latency for these messages can be achieved.

In normal operation reconfiguration would occur only in the case of component failure, battle damage, or change of mission. A reconfiguration may be initiated by any node by sending a request to the node controlling the crossbar switch. If the request is valid, this node instructs the interface circuits at the switch ports to begin issuing reset commands around the affected ringlets. The crossbar switch is then set and the affected ringlets are allowed to reinitialize in the standard way. During initialization new node IDs are assigned to each node if necessary. The entire process is estimated to take less than 1 ms. In comparison, the SAE requirement for reconfiguration of an SDDN is 50 ms [SAE93].

The current prototype operates at a serial data rate of 1-Gbps. This rate is limited by the speed of the crossbar switch. If a faster electrical or optical switch were available the ultimate speed of the switch would be 1.6-Gbps, limited by the speed of the interface circuitry. The back-end bus operates at an aggregate data rate of 3.2-Gbps.

The power dissipation of the switch may be estimated from the individual components. Each port consists of an optical transceiver, a serializer/deserializer, interface circuit, and assorted line drivers. Total power dissipation for these components is 11.15 W. In addition, the crossbar switch and control logic dissipate 5.4 W. Total power dissipation for the 5-port prototype is estimated to be 61.15 W. A 16-port version would dissipate 183.8 W.

Simulation Descriptions

The following sections provide descriptions of the models that were created to simulate the SCI protocol and different SCI switches to measure the performance of complete systems. The SCI emulation model provides the basic SCI transport operations in a fine-grain manner. The switch models extend the emulation model to simulate a packet-level switch as well as the MSS. Three example systems are presented and network loading scenarios are described to show the relative benefits of each of the topologies. Finally, results of the simulations are presented and analyzed.

SCI Emulation Model

The SCI emulation model was designed and implemented using the Block-Oriented Network Simulator (BONeS) from the Alta Group of Cadence Systems, Inc. BONeS is a discrete-event simulator with many built-in modeling blocks for fine-grain network simulation. The SCI emulation model was designed to follow the SCI standard as closely as possible, sacrificing minimal fidelity to improve simulation speed. The model has many parameters that can be set to match experimental measurements of existing SCI hardware. In this way, specific hardware implementations can be simulated by calibrating the model using these parameters.

The model was built to be generic and reusable although some design parameters were assumed. First, packet routing is of prime importance when modeling any switches. The SCI node routing decisions are made by table lookups of routing tables which are dynamic and can be rewritten during simulation if reconfiguration occurs. Generic routing tables can also simulate static-routing schemes such as interval routing. A symbol-level simulation is most desirable for fidelity purposes but can lead to extremely long simulation times. Instead, two modeling techniques were used to improve simulation time. First, any output symbols of a contiguous SCI packet are clumped together. In this way, only one event is triggered once a packet is received instead of the 40 events for a 40- symbol send packet. Second, the packet undergoes a "pipelined" delay during reception. This technique forces the receiving node to delay until the needed symbol of the packet arrives before it is allowed to use the information. In this way, exact bypass and routing delays can be simulated with great accuracy.

Each node has an adjustable clock frequency and is assumed to output a single 18-bit symbol during each clock period. Hence, serial SCI nodes can be simulated by appropriate clock frequency selections. The node's host interface is separately clocked to simulate a different speed host. The host interface was designed to support either an asynchronous or synchronous host. An asynchronous host offers traffic at an arbitrary rate and will process rejected packets if the output queue is full. An asynchronous host will attempt to service the input queue as quickly as possible. If the host is not available, the host rejects the incoming packet which is pushed back into the input queue. If the host cannot service incoming packets at a sufficient rate, the input queue will fill which forces new packets to be retried using SCI's queue reservation protocols for retried packets.

Synchronous hosts offer packets at a constant rate to the output queue and service packets at a constant rate from the input queue. This mode of packet handling simulates constant rate sources such as sampling sensors and constantly-polled input sinks. The modeled interface was designed in such a way to support both timing methods simultaneously.

SCI Switch Models

A packet switch is shown in Figure 4 and is built of multiple SCI nodes. The host interfaces of the nodes in the switch are connected to a common fabric such as a shared bus.

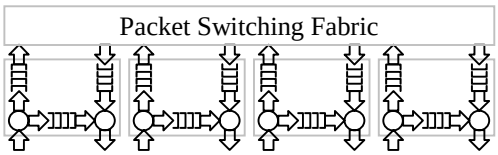

Figure 4: SCI Packet Switch

The MSS is built by combining a packet switch with a crossbar to allow switching of physical circuits. This design is shown in Figure 5. Notice that once rings are combined using the crossbar, the SCI nodes inside the switch simply pass packets destined for a node on the new ring through their bypass FIFOs instead of stripping them off and passing them over the packet switching fabric.

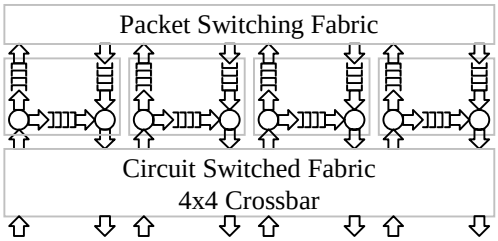

Figure 5: SCI Multiservice Switch

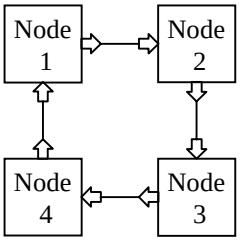

Figure 6: 4-Node SCI Ringlet System

Simulated Systems

A simple SCI ringlet system, shown in Figure 6, was used as a baseline for comparisons of latency, throughput, and response time variance. The ringlet is formed by connecting the output link of one node to the input link of the following node and requires no additional hardware. A system of 4 nodes connected with a packet switch was used to verify the routing performance of the switch. The packet switch system is shown in Figure 7. This configuration offers a separate ringlet per node and requires a high-performance, packet-switching fabric to maintain high throughput. Finally, a system built with an MSS is shown in Figure 8. The configuration is isomorphous to Figure 7, as the MSS is topologically identical to a packet switch.

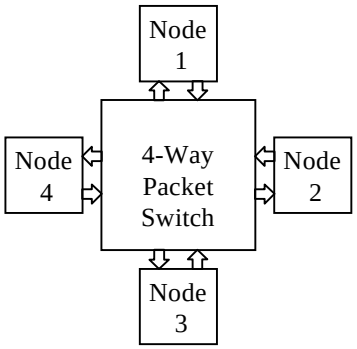

Figure 7: Connectionless Switch System

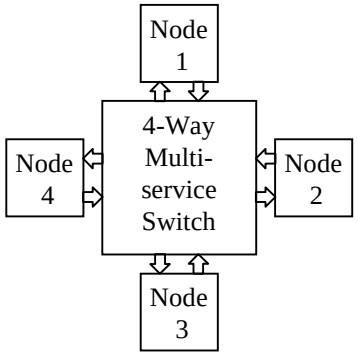

Figure 8: Multiservice Switch System

Simulation Scenarios

In order to gauge the effectiveness of the multiservice switch, the three systems described above were implemented in the simulation environment. Each node of the system was configured with a statistical requester and an active responder. The requester has four types of parameters that can be varied to simulate certain classes of data sources: request type, interarrival type, burst type, and destination type.

- The *request type* specifies which commands this requester will generate and at what size. Common commands are read, write, and move with standard payload sizes of 64 or 256 bytes per request. A read command requests a certain block of memory from the responder, which generates a response packet with the data. A write command passes a block of data to the responder to write into memory. The responder replies with a response packet once the data has been committed into memory. A move command writes data from the requester to the responder but eliminates the response subaction.
- The *interarrival type* specifies the time between subsequent requests. Available interarrival rates are fixed or random with uniform, exponential, or normal distributions. The mean and variance can be specified.
- The *burst type* specifies how many requests are generated in a stream from this requester. The number of requests can be fixed or random with uniform, exponential, or normal distribution, again with mean and variance as parameters.
- The *destination type* specifies where requests from this node will be sent. The available destinations are fixed, random with uniform distribution, downstream (next node on ring), upstream (previous node on ring), and self.

By selecting the appropriate parameters of the source, different loading conditions on the network can be investigating in hopes to predict actual performance. Parameters that specify SCI node performance can also be varied and reasonable choices were chosen. Table 1 lists the externally-variable node parameters and the values chosen throughout all simulations.

Table 1: SCI Simulation Parameters

Parameter	Description	Value
Input Queue Size	Number of packets that can be stored in the input queue	3
Output Queue Size	Number of outstanding transactions	3
Link Data Rate	Speed that raw data is passed over SCI	1.6 Gbps (i.e. 200 MBps)
Host Data Rate	Speed that raw data is passed from the SCI node to the host	1.6 Gbps (i.e. 200 MBps)
Switch Data Rate	Speed that raw data is passed through the packet switching fabric	3.2 Gbps (i.e. 400 MBps)
Stripping Delay	Symbols necessary to determine packet destination, w/ no routing table check	2 symbols
Routing Table Delay	Symbols necessary to delay while checking the routing table	40 symbols (store and forward switches)
Link Length	Length of electrical wiring runs between nodes	3 meters

Each of the three network configurations was offered the three following loading conditions to allow a fair comparison between the topologies. Table 2 summarizes in qualitative terms the expected results of the simulation.

1. The first loading condition is a streaming test. This involves two nodes (the first and the fourth) in which node 1 sends 64-byte move packets to node 4 at a fixed rate. The throughput and latency is calculated at the responder node. This test forms the upper bound in throughput for the specific topology. The switched MSS system performance is expected to match the ringlet system while the packet switched system will have a slight decrease in throughput due to routing delays.
2. The second loading condition offers a varying total offered load to each system where each node sends a fixed burst length of read and write requests to a random responder with a Poisson distributed interarrival rate. The latency and throughput is measured at the requester since reads and writes are response-expected transactions. The ring

performance is expected to be poor since the ring bandwidth is fairly shared among all 4 nodes. The two switches are expected to perform identically since the MSS gains no advantage of circuit switching under random traffic. The switched systems will enjoy a much higher aggregate throughput than the ring system due to the separated ringlets.

3. The final loading condition combines the first two to mimic a typical avionics sensor-processing workload. Node 1 is specified as a source node and streams data to node 4. Simultaneously, all nodes except node 1 send out fixed burst messages to random destinations. The streaming load is made up of 64-byte move transactions and is representative of sampled data from a sensor. The random load is typical of control messages and uses an exponential interarrival rate to simulate computer-generated traffic. The streaming data is designed to utilize 10 times the bandwidth of the combined random load. Actual SAE specifications cite streaming loads up to 2-Gbps and control loads up to 1-MBps, a 200:1 ratio [SAE93]. In this final case, the MSS should show the streaming performance of the ring and the bursty performance of a switch while the packet switched system and the ring system will perform worse due to topological constraints.

Table 2: Qualitative Expected Results

	Streaming	Random	Mixed
Ring	Good	Poor	Poor
Packet Switch	Poor	Good	Poor
MSS	Good	Good	Good

Simulation Results

The simulation results are grouped into sections based on the three loading conditions. The first set of graphs shows the throughput and latency for the streaming-load scenario.

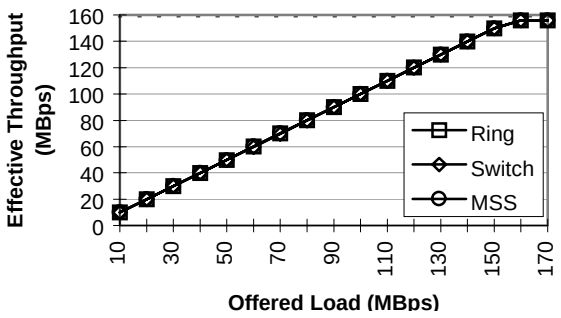

Figure 9: Streaming Scenario Throughput Results

Figure 9 shows that all three topologies can handle a single source saturating the network and all three saturate at the same rate (i.e. 160 MBps, which is 40 MBps less than the link data rate due to packet overhead). This chart does not show how much bandwidth is available after the network saturates. Since the ring topology shares bandwidth, very little bandwidth is available with a single high-load source. Both of the switch systems still have full bandwidth available on ringlets 2 and 3. The packet switch system has half of the internal fabric bandwidth remaining while the MSS has the full internal fabric bandwidth remaining.

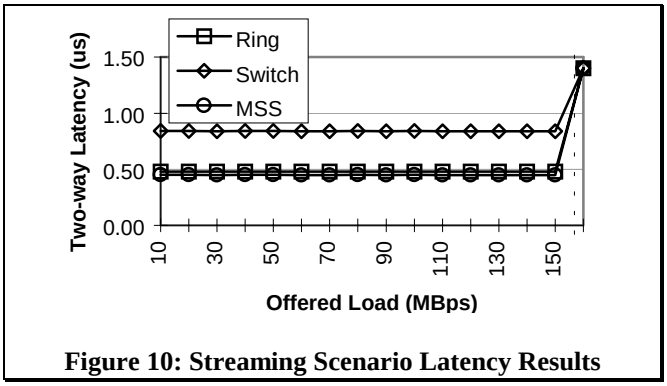

Figure 10: Streaming Scenario Latency Results

Figure 10 shows the latency for the streaming-load scenario. The packet switch system has a fundamentally higher latency than both the ring and MSS systems. This is the routing delay. Both the MSS and ring avoid any packet switching and therefore enjoy a lower minimum latency by approximately 0.4 µs. The MSS has a slightly lower latency than the ring due to the configuration. The number shown is the two-way latency of packets that were actually received. In the overloading case, latency is infinite since some packets will never reach their destinations so an appropriate number was chosen for display purposes.

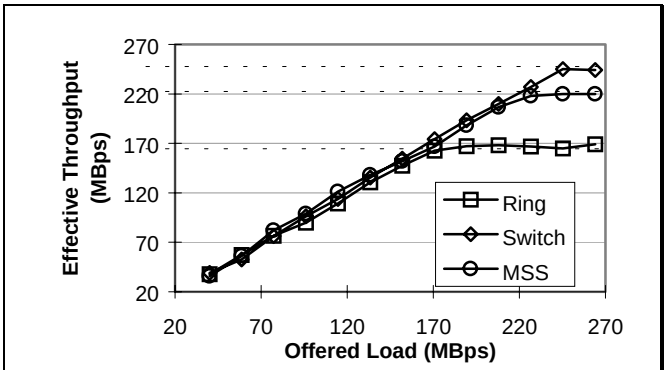

Figure 11: Random Scenario Throughput Results

Figure 11 shows the throughput of the random load scenario. This scenario shows the benefit of using a switched topology. Notice how the saturation bandwidth of both switched systems is higher than the ring which saturates at 166 MBps. The MSS, which has nodes 1 and 4 circuit switched onto the same ringlet, has a higher bandwidth than the ring due to its packet switch fabric but has a smaller throughput than the switched system due to the circuit-switched ringlet. Here, approximately half of the load uses the ring while half uses the packet switching (due to uniform distribution of destinations). Hence the performance of the MSS system is about halfway between the packet switched system and the ring system.

Figure 12 shows the latency for the random destination loading scenario. A distinction between the three systems can be seen here. Again, the performance of the MSS system is approximately halfway between the ring system and the packet-switched system. The packet-switched network has the lowest average latency for the random destination case. This occurs due to the sharing of bandwidth on the ring system as well as the ringlet in the MSS system.

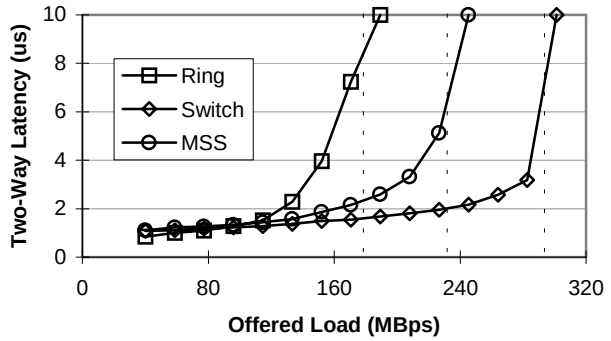

Figure 12: Random Scenario Latency Results

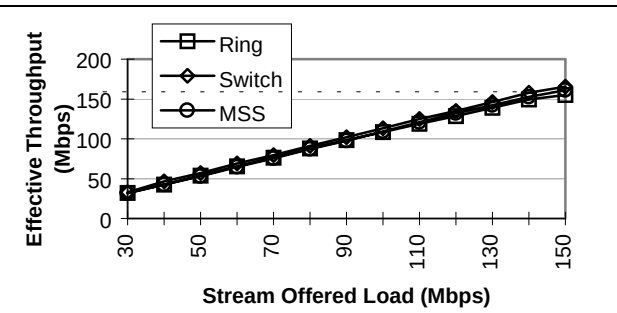

Figure 13: Mixed Scenario Throughput Results

Figure 13 shows the mixed load throughput results. Again, all three systems are able to saturate the network at the streaming load limit of 150 MBps. Recall that the mixed load is composed of the streaming load from node 1 and the random destination load that is $1/10^{th}$ the streaming load (i.e. nodes 2,3, and 4 transmit at $1/30^{th}$ the rate as node 1).

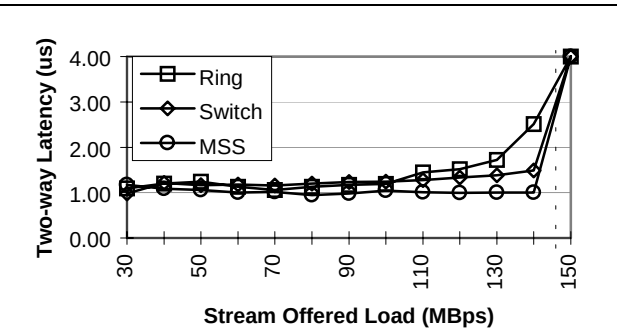

Figure 14: Mixed Load Latency Results

Figure 14 shows the latency of the mixed load scenario. Here, only two-way latency of the random destination packets for comparison with the random destination test. Under the mixed load scenario, the MSS system maintains the lowest average latency for the random destination packets while also having throughput that is as equally high as the other topologies.

Conclusions

This paper presented the design, modeling, and simulation of a novel switching technique for next generation avionics data networks. The multiservice switch offers two switching mechanisms to gain the performance and fault-tolerance benefits of a packet switch while simultaneously offering the low latency of a ring-based topology.

The performance improvements of the multiservice switch will allow system designers to reduce the packet switch speed requirements to attain the same level of performance for streaming loads. By reducing the speed of the packet switch, power and cost are reduced. The multiservice switch also shows equal if not better performance than conventional switches and topologies for mixed offered loads, which can be expected in an avionics data network.

Future Research

Future work on the multiservice switch will complete the prototype switch in both hardware and software. The prototype switch still requires control software to be written and some hardware debugging. The simulator will be expanded to handle actual, rather than statistical, offered loads and to include more efficient switching mechanisms. The simulator will also be expanded to simulate the actions necessary for a run-time crossbar reconfiguration.

Acknowledgments

We would like to thank the Office of Naval Research for their support of this research as well as Michael Miars and Robert Todd of the High-performance Computing and Simulation Research Lab for their technical assistance.

References

[UHLH92] R.W. Uhlhorn, C.E. Hudgins, and T.A. McDermott, "Very High Speed Optical Networks (VHSON) Final Report for Period March 1990 - April 1992," Wright Laboratory Document Number WL-TR-92-1065, May 1992, p. 59.

[SAE93] Society of Automotive Engineers, "High Speed Data Transfer Networks Requirements Subgroup Report", Rev. 1.2, February 18, 1993.

[JAST94] Joint Advanced Strike Technology Program, "Avionics Architecture Definition," Ver. 1.0, September 23, 1994.

[SCI93] Institute of Electrical and Electronics Engineers, "Standard 1596-1992 for the Scalable Coherent Interface (SCI)", New York, NY, 1993.